

## **SYSTEM AND METHOD FOR SELF-DIAGNOSING SYSTEM CRASHES**

### **FIELD OF THE INVENTION**

The present invention relates to computer system crash  
5 analysis. More specifically, the invention relates to the  
identification of a component responsible for a computer system  
crash.

### **BACKGROUND OF THE INVENTION**

Today, diagnosing a computer system crash (due to  
10 operating system or device driver software bugs, hardware  
errors, configuration problems, or the like) is a very time  
consuming and expensive process. Typically, a system  
administrator or developer is left to access books, websites,  
or colleagues, and often resorts to trial and error to  
15 determine what exactly caused the system crash. The diagnosis  
is generally manual and involves setting particular diagnostic  
configurations, rebooting the system (likely many times),  
manually evaluating the diagnostic results, and attempting to  
reproduce the crash.

20 In some operating systems, when a crash occurs, a dump  
file may capture the operating state of the computer at the  
time of the crash. The traditional dump file helps solve the  
mystery of what caused the crash, but is typically a very large  
file. For instance, large systems may have several gigabytes  
25 of memory. Writing out the traditional dump file may take

upwards of thirty minutes on such a system. Users typically disdain that much down time, and administrators prefer to avoid such time-consuming steps toward diagnosing the system crash.

Moreover, as suggested above, using the information stored  
5 in the dump file has traditionally been a time-intensive, manual process. A system administrator or developer is left to read many lines of information in an attempt to determine what caused the crash. Hours of human intervention may be spent simply identifying the diagnostic steps to be taken in search  
10 of the offending component that caused the crash.

Further complicating the diagnosis of system crashes is that they are often difficult to reproduce. For example, a device driver may have a bug that does not arise unless memory is low, and then possibly only intermittently. In that case, a  
15 test system may not be able to reproduce the error because it does not reproduce the conditions.

In sum, diagnosing system crashes has long vexed system administrators and users of computing systems. A system that overcomes the problems identified above has eluded those  
20 skilled in the art.

### **SUMMARY OF THE INVENTION**

Briefly described, the present invention provides a system and method for self-diagnosing system crashes by identifying a  
25 type of system crash that occurred, and automatically taking

diagnostic steps based on that type of crash. The invention may make use of a stop code contained in a memory dump file stored in response to the system crash. Preferably, the invention makes use of a "minidump" that contains an

5 intelligently selected subset of the available pre-crash computer information, including the stop code that identifies the particular type of crash that occurred.

In one implementation, a mechanism of an operating system is configured to write an abbreviated dump file of a selected  
10 portion of the system memory at the time of a system crash. For example, a "crash driver" may be implemented that, when instructed by the operating system, reads from system memory certain information considered to be likely the most relevant to the diagnosis of a system crash, and writes that information  
15 to the dump file. Typically, a component of the operating system (e.g., a memory manager component) identifies the occurrence of a system fault, such as corrupt or exhausted memory, and informs the operating system that the system crash has occurred. In response, the crash driver may be instructed  
20 to write the dump file so that the crash may be diagnosed.

In accordance with an aspect of the invention, another mechanism within an operating system, such as a memory management component of a system kernel, checks for the existence of the dump file at each startup of the machine. The  
25 existence of the dump file may indicate that the system crashed

during the previous session. The existence of the dump file is but one technique that may be used to determine that a system crash occurred, and is only given as an example. In any case, once the occurrence of the system crash has been discovered, the mechanism of the invention analyzes the dump file to determine what type of crash occurred (e.g., out of memory or corrupt memory), and implements a self-diagnostic routine or procedure corresponding to the type of crash. More particularly, the mechanism may read the stop code from the dump file and implement a self-diagnostic procedure that corresponds to that stop code.

Through the described construct, the mechanism self-diagnoses the likely cause of the crash by automating many of the tasks normally performed manually. If the crash occurs again, the mechanism identifies, through the self-diagnostic procedures automatically implemented, the likely cause of the crash, e.g. the particular faulty driver or configuration error, and may report that information to a system administrator. This significantly simplifies the corrective measures that typically need to be taken by system administrator or the like to correct the fault. Moreover, the self-diagnostic procedure may enable special code to provoke the problem into reoccurring sooner, and, more importantly, to also catch it before it causes too much damage so the culprit can be easily identified. And still further, the invention

enables non-experts to quickly diagnose and resolve computer problems, thereby ameliorating both the cost and delay of finding an "expert."

In accordance with another aspect of the invention, during  
5 startup, the mechanism may change the stop code stored in the dump file to avoid a situation where the system suffers another, different type of crash before the mechanism is able to address the first crash (such as later in the startup process).

10 These and other aspects of the invention, together with the benefits and advantages realized, will become apparent from a reading of the following detailed description in conjunction with the drawings, in which:

15 **BRIEF DESCRIPTION OF THE DRAWINGS**

Figure 1 is a block diagram representing a computer system into which the present invention may be incorporated;

Figure 2 is a functional block diagram representing a self-diagnostic system capable of being used in conjunction  
20 with the computer system of Figure 1 and in which the present invention may be embodied;

Figure 3 is a functional block diagram illustrating components of the self-diagnostic system used to detect and identify a system crash, and to store information to self-

diagnose the cause of the system crash, in accordance with one aspect of the present invention;

Figure 4 is a logical flow diagram illustrating a process performed by the system illustrated in Figure 3 to detect and  
5 identify the system crash, and to store information to self-diagnose the cause of the system crash, in accordance with one aspect of the present invention;

Figure 5 is a functional block diagram illustrating components of the self-diagnostic system configured to detect  
10 the occurrence of a previous system crash, and to self-diagnose the cause of the system crash, in accordance with one aspect of the present invention;

Figures 6A and 6B are logical flow diagrams generally illustrating a process performed by the components of Figure 5  
15 to detect and self-diagnose a system crash, in accordance with one aspect of the invention;

Figure 7 is a representation of data structures uses to track memory allocations and deallocations to diagnose a particular type of system crash;

20 Figure 8 is a functional block diagram illustrating a type of system crash resulting from exhaustion of virtual memory, and components of the self-diagnostic system used to diagnose it; and

Figure 9 is a functional block diagram illustrating a type  
25 of system crash resulting from exhaustion of virtual memory

resources, and components of the self-diagnostic system used to diagnose it;

Figure 10 is a functional block diagram illustrating a particular type of system crash resulting from locked pages not  
5 being released, and components of the self-diagnostic system used to diagnose it; and

Figures 11 and 12 are functional block diagrams illustrating components of the self-diagnostic system used to diagnose various types of system crashes resulting from corrupt  
10 memory.

### **DETAILED DESCRIPTION**

#### **EXEMPLARY OPERATING ENVIRONMENT**

Figure 1 and the following discussion are intended to  
15 provide a brief general description of a suitable computing environment in which the invention may be implemented.

Although not required, the invention will be described in the general context of computer-executable instructions, such as program modules, being executed by a personal computer.

20 Generally, program modules include routines, programs, objects, components, data structures and the like that perform particular tasks or implement particular abstract data types.

Moreover, those skilled in the art will appreciate that the invention may be practiced with other computer system  
25 configurations, including hand-held devices, multi-processor

systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers and the like. The invention may also be practiced in distributed computing environments where tasks are performed by  
5 remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

With reference to Figure 1, an exemplary system for  
10 implementing the invention includes a general purpose computing device in the form of a conventional personal computer 20 or the like, including a processing unit 21, a system memory 22, and a system bus 23 that couples various system components including the system memory to the processing unit 21. The  
15 system bus 23 may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures. The system memory includes read-only memory (ROM) 24 and random access memory (RAM) 25. A basic input/output system 26 (BIOS),  
20 containing the basic routines that help to transfer information between elements within the personal computer 20, such as during start-up, is stored in ROM 24. The personal computer 20 may further include a hard disk drive 27 for reading from and writing to a hard disk, not shown, a magnetic disk drive 28 for  
25 reading from or writing to a removable magnetic disk 29, and an



optical disk drive 30 for reading from or writing to a removable optical disk 31 such as a CD-ROM, DVD-ROM or other optical media. The hard disk drive 27, magnetic disk drive 28, and optical disk drive 30 are connected to the system bus 23 by a hard disk drive interface 32, a magnetic disk drive interface 33, and an optical drive interface 34, respectively. The drives and their associated computer-readable media provide non-volatile storage of computer readable instructions, data structures, program modules and other data for the personal computer 20. Although the exemplary environment described herein employs a hard disk, a removable magnetic disk 29 and a removable optical disk 31, it should be appreciated by those skilled in the art that other types of computer readable media that can store data that is accessible by a computer, such as magnetic cassettes, flash memory cards, digital video disks, Bernoulli cartridges, random access memories (RAMs), read-only memories (ROMs) and the like may also be used in the exemplary operating environment.

A number of program modules may be stored on the hard disk, magnetic disk 29, optical disk 31, ROM 24 or RAM 25, including an operating system 35 (preferably the Windows® 2000 operating system). The computer 20 includes a file system 36 associated with or included within the operating system 35, such as the Windows NT® File System (NTFS), one or more application programs 37, other program modules 38 and program

data 39. A user may enter commands and information into the personal computer 20 through input devices such as a keyboard 40 and pointing device 42. Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish,  
5 scanner or the like. These and other input devices are often connected to the processing unit 21 through a serial port interface 46 that is coupled to the system bus, but may be connected by other interfaces, such as a parallel port, game port or universal serial bus (USB). A monitor 47 or other type  
10 of display device is also connected to the system bus 23 via an interface, such as a video adapter 48. In addition to the monitor 47, personal computers typically include other peripheral output devices (not shown), such as speakers and printers.

15 The personal computer 20 may operate in a networked environment using logical connections to one or more remote computers 49. The remote computer (or computers) 49 may be another personal computer, a server, a router, a network PC, a peer device or other common network node, and typically  
20 includes many or all of the elements described above relative to the personal computer 20, although only a memory storage device 50 has been illustrated in Figure 1. The logical connections depicted in Figure 1 include a local area network (LAN) 51 and a wide area network (WAN) 52. Such networking

environments are commonplace in offices, enterprise-wide computer networks, Intranets and the Internet.

When used in a LAN networking environment, the personal computer 20 is connected to the local network 51 through a network interface or adapter 53. When used in a WAN networking environment, the personal computer 20 typically includes a modem 54 or other means for establishing communications over the wide area network 52, such as the Internet. The modem 54, which may be internal or external, is connected to the system bus 23 via the serial port interface 46. In a networked environment, program modules depicted relative to the personal computer 20, or portions thereof, may be stored in the remote memory storage device. It will be appreciated that the network connections shown are exemplary and other means of establishing a communications link between the computers may be used.

#### ILLUSTRATIVE SELF-DIAGNOSTIC SYSTEM

Figure 2 is a functional block diagram that, together with the following discussion, provides a general description of an illustrative self-diagnostic system 200 constructed in accordance with the present invention. Components of an illustrated embodiment of the system 200 include a memory manager 201, which may be a kernel-mode component of the operating system 35, a crash driver 203, and a boot loader 205. Also illustrated are one or more kernel-mode device drivers

207, a non-volatile storage device, such as the hard disk 27,  
and several pages of the system memory 22. The several  
components of the system 200 are illustrated as separate from  
the system memory 22 for clarity of explanation, but those  
5 skilled in the art will appreciate that those components may  
reside within the system memory 22 during execution.

The memory manager 201 of the described embodiment is a  
kernel-mode operating system component that (among other  
things) implements on-demand-paged virtual memory. The memory  
10 manager 201 allocates and deallocates memory, pages data into  
and out of physical memory (from and to a disk) as memory  
demands dictate, and generally maintains the flow of data into  
and out of system memory 22. The memory manager 201 performs  
such tasks to make it appear to other software components, such  
15 as application programs 37 or the drivers 207, that there is  
more memory available (e.g., four gigabytes of virtual memory)  
than the physical random access memory actually installed on  
the computer 20 (e.g., 128 megabytes of RAM). For example,  
when a driver needs memory, it requests a certain amount from  
20 the memory manager 201, which then allocates the memory (if  
available) to the driver. The memory manager 201 may trim data  
stored in the allocated memory to hard disk if not actively  
used by the driver, thereby freeing the physical memory for use  
by another component until again used by the driver.

The crash driver 203 is a special type of system driver that is responsible for writing a dump file to the hard disk 27 in the event of a system crash. More specifically, if the memory manager 201, or some other component of the operating system 35, identifies a system memory fault, that component instructs the crash driver 203 to read selected data from memory 22 and write that data to the hard disk 27 as the dump file. The crash driver 203 is provided a stop code by the operating system 35 that indicates a particular type of fault that led to the system crash. For example, an out-of-memory fault may correspond to one stop code, while a corrupt-memory fault may correspond to another stop code. The component that identifies the system memory fault may provide the stop code to the crash driver 203. Although briefly described here, further discussion of the crash driver 203 and the method of creating the dump file can be found in copending U.S. Patent Application Number \_\_\_\_\_, filed on April 14, 2000, entitled "Methods and Arrangements for Generating Debugging Information Following Software Failures," assigned to the same assignee as the present patent application, and incorporated herein by reference.

The boot loader 205 is a component of the operating system 35 that contains the code used to load the operating system 35 into RAM 25 at system start up. In one embodiment, the boot loader 205 is configured to query for the existence of the dump

file in a particular location on the hard disk 27. For instance, if the crash driver 203 is configured to write the dump file to a pagefile on the hard disk 27, the boot loader 205 is configured to read the dump file from the pagefile at start up. In this embodiment, the existence of the dump file indicates to the boot loader 205 that the system crashed during the previous session. Other mechanisms may alternatively be used to indicate the occurrence of the system crash, such as a status bit or code that is tested at each start up, or any other technique. The boot loader 205 is further configured to extract the stop code and other pertinent information from the dump file and pass it to the memory manager 201 early in the boot process (i.e., prior to loading any device drivers 207). In that way, the memory manager 201 knows the type of crash that occurred and may take action (as described below) prior to starting any drivers.

The device drivers 207 are software components that make use of the system memory 22 to perform varied tasks associated with the operation of the computer system 20. Often device drivers are associated with hardware devices, such as a hard disk driver or sound card driver. However, some device drivers perform specialized functions and are not associated with a particular hardware device, such as an antivirus driver or a file encryption driver. The device drivers 207 interface with the memory manager 207 to request the allocation, deallocation,

and general maintenance of memory 22. Very often device drivers are kernel mode components that execute with a privileged status, which provides them with certain access privileges to otherwise protected memory that results in improved efficiency and performance. However, an ill-behaved device driver executing with privileged status can cause system crashes by improperly accessing or maintaining memory 22. Often, the memory manager 201 can detect that an errant device driver caused a memory fault, but cannot detect specifically which device driver 207 caused the fault, leading to the need to diagnose the system crash to identify the errant driver.

Figure 3 and Figure 4 are a functional block diagram and a logical flow diagram, respectively, that together generally illustrate how the system 200 may detect and identify a system crash, and store information used by the system 200 to self-diagnose the cause of the system crash. To begin, as represented by block 401 of Figure 4, the system 200 performs normally, meaning that ordinary memory maintenance and accesses occur, until a system crash is detected by the operating system 35, such as by the memory manager 201. When a system crash is detected (e.g., a memory fault), at block 403 the memory manager 201 identifies the type of crash that occurred and assigns a stop code 311 (Figure 3) to that type. For instance, if the memory manager 201 detects a dangerously-low memory condition resulting from a device driver failing to release

unnneeded memory allocations, the memory manager 201 identifies the particular stop code 311 associated with that crash type. The memory manager 201 may refer to a lookup table or other listing that maps stop codes to memory faults or other crash  
5 types.

At block 405, the memory manager 201 indicates the occurrence of a system crash to the crash driver 203, such as by passing the stop code 311 to the crash driver 203. Other techniques may equally be used to inform the crash driver 203  
10 of the crash.

At block 407, once the crash driver 203 has the stop code 311 and has been informed of the crash, the crash driver 203 reads certain data from memory 22 to capture the state of the system at the time of the crash. The data read from memory 22  
15 is selected to allow effective analysis and diagnosis of the crash. Further detail on the selected data and the operation of the crash driver 203 may be found in the aforementioned United States Patent Application Number \_\_\_\_\_.

At block 409, the crash driver 203 writes the dump file  
20 315 to the hard disk 27. In an embodiment described herein, the crash driver 203 writes the dump file 315 to a particular location on the hard disk 27, such as a pagefile or other known location on the hard disk 27. In that way, the existence of the dump file 315 may be easily verified during a subsequent  
25 start up operation. The stop code 311 is included within the



dump file 315 when written to the hard disk 27. Alternatively, the stop code 311 may be written to the hard disk 27 or pagefile separately from the dump file 315.

At block 411, after the stop code 311 has been written to the hard disk 27, the operating system 35 may shut down the computer system 20 or otherwise notify the user of the system crash. In some situations, the operating system 35 may be able to avoid completely shutting down the computer system 20 despite the crash, such as by terminating certain executing applications or processes that are affected by the crash. In other situations, the operating system 35 will prompt for (or otherwise cause) a reboot of the system, such as the case where operating system information has been corrupted by the crash. At this juncture, the dump file 315 and the stop code 311 have been saved to the hard disk 27, and the system 200 is prepared to self-diagnose the cause of the crash in accordance with one aspect of the present invention.

Figure 5 is a functional block diagram and Figures 6A and 6B are logical flow diagrams that, taken together, generally describe how the system 200 determines that self-diagnosis is appropriate, and how the system 200 performs the self diagnosis. As represented in Figure 5, during system boot, the boot loader 205 is launched to load the operating system 35 into memory 22. In an embodiment of the invention described herein, at block 603, the boot loader 205 checks for the

existence of the dump file 315 on the hard disk 27 to determine whether the system crashed during the previous session. As mentioned above, the dump file 315 may be stored in a page file on the hard disk 27 so that it may be easily located by the  
5 boot loader 205 at start up.

At decision block 605, a test is made to determine whether the dump file exists. If the dump file does not exist, the process illustrated in figure 6A proceeds to block 607, where the system continues to boot normally, and the boot process  
10 effectively terminates at ending block 609. If however, the dump file 315 does exist, the system suffered a crash during the previous session, and the process illustrated in Figure 6A continues to block 611.

At block 611, the boot loader 205 retrieves the dump file  
15 315 from the hard disk 27. As discussed above, the dump file 315 may reside in a pagefile or some other known location on the hard disk 27 for easy location by the boot loader 205. At block 613, the boot loader 205 retrieves the stop code 311 from the dump file 315, and, at block 615, passes the stop code 311  
20 and the contents of the dump file 315 to the memory manager 201.

The process continues at decision block 621 illustrated in Figure 6B, where the memory manager 201 determines if a corresponding diagnostic procedure exists for the stop code  
25 311. Note that all possible system crashes for which a stop

code exists may not have a corresponding diagnostic procedure. Thus, if the stop code 311 does not have a corresponding diagnostic procedure, the memory manager 201 may either execute a default diagnostic procedure (as illustrated by "Yes" branch  
5 of the decision block 623), or proceed to boot normally at block 625. If a default diagnostic procedure is executed, the process may continue to block 629.

If at block 621 a corresponding diagnostic procedure exists for the stop code 311, at block 627, the memory manager  
10 201 executes the diagnostic procedure for the stop code 311. The diagnostic procedure may include first ensuring that the computer system, including any device drivers and configuration settings, is still in the same configuration as before the crash, in case an administrator or other person identified and  
15 remedied the cause of the crash prior to the first post-crash reboot. The memory manager may verify this by comparing the list of components in the dump file 315 with those in the currently-running system. If the system is still in the same configuration, the memory manager 201 modifies certain system  
20 variables (identified by the diagnostic procedure) to enable appropriate switches (e.g., debug/tracking/sandboxing/driver verification switches) and the like, for the particular stop code 311. With any appropriate system variables set, the particular driver that caused the previous crash likely will be

identifiable upon the next crash because it will typically be caught by these safeguards much sooner, as described below.

At block 629, the system continues the start-up process until the operating system 35, the device drivers 207, and any  
5 other appropriate software modules have been loaded into system memory 22. At this juncture, the computer system 20 is ready for use, and the diagnostic system 200 is prepared to identify the cause of the previous system crash in the event that it reoccurs.

10 At block 631, the memory manager 201 monitors the system operation in accordance with the system variables set at block 627. For example, the memory manager 201 may maintain an allocation table 511 that tracks particular types of memory usage by the device drivers 207. To this end, as generally  
15 represented in Figure 7, the memory manager 201 may maintain a verification block (e.g., 701<sub>1</sub>-701<sub>n</sub>) that points to an outstanding allocation table (e.g., 511<sub>1</sub>-511<sub>n</sub>) for each driver that requests or performs a particular type of memory operation if the previous crash resulted from that type of memory  
20 operation (as identified by the crash code 311). As generally represented in FIG. 7, the verification block (e.g., 701<sub>1</sub>) comprises driver information including a count of the total allocations for this driver. Other information may be  
25 allocation information for this driver. The verification block

701<sub>1</sub> also includes a pointer to the outstanding allocations table 511<sub>1</sub> set up for a corresponding driver. The outstanding allocations table 511<sub>1</sub> tracks specific information about each pool allocation that the driver has been given that remains  
5 outstanding, i.e., has not yet deallocated. The information may include the allocation's virtual address, length, and information useful in debugging such as per-process caller information and the tag of the driver that allocated the memory, e.g., "TCP" for TCP/IP drivers.

10        Thus, for each memory allocation to a driver, the memory manager 201 may include an entry in the allocation table (e.g., 511<sub>1</sub>) associated with that driver that requests or performs a particular type of memory operation if the previous crash resulted from that type of memory operation (as identified by  
15 the crash code 311). In that way, when the crash reoccurs, the allocation tables 511<sub>1</sub>-511<sub>n</sub> will likely identify which driver, drivers, or other software program being tracked is the offending culprit. Note that it is alternatively feasible to determine from the various counts if a given driver is  
20 consuming inordinate amounts of memory.

      The system continues operating and the memory manager 201 continues monitoring the system operation until, illustrated by decision block 633, the system crash reoccurs and the process proceeds to block 635. When the system crashes again, the  
25 process illustrated in Figure 6B moves to block 635. At block

635, the diagnostic mechanisms employed by the memory manager  
201 contain an identification of each driver or software module  
that has outstanding memory operations of the type identified  
by the stop code 311. For example, if the stop code 311  
5 indicates that the previous system crash occurred as a result  
of low memory, the allocation table 511 may indicate the  
drivers that had allocations and details about those  
allocations at the time of the crash. Thus, the memory manager  
201 reports the results of the diagnostic procedure to the user  
10 or system administrator in the form of a printout or summary of  
the allocation tables, or any other form of report. In this  
manner, the user or system administrator may evaluate the  
results to help determine which driver or software component in  
the computer system caused the crash. For instance, in the  
15 illustrated example, one of the drivers may be associated with  
a much larger number of allocated pages at the time of the  
second crash relative to other drivers. That information  
indicates to the user or system administrator that that driver  
is likely the cause of the crash, and should be perhaps  
20 replaced or further tested. Thus, the system 200 is able to  
self-diagnose the cause of the previous system crash, and upon  
a subsequent crash, provide the user or system administrator  
with an identification of the likely culprit behind the crash:

To help further clarify the operation of the disclosed  
25 self-diagnostic system, the following discussion provides

several specific examples of types of system crashes that may occur, and how the system 200 illustrated above may self-diagnose those types of crashes. The examples fall into two broad categories, out-of-memory/resources crashes and corrupt-  
5 memory crashes.

#### OUT-OF-MEMORY/RESOURCES SYSTEM CRASHES

The first category of system crashes deals with a failure of a software component, such as a device driver, to deallocate  
10 memory resources that the component no longer needs, often referred to as a "memory leak." Memory leaks can occur, for example, if a driver unloads but still has allocated memory, or even when a driver is loaded but for some reason does not deallocate unneeded memory. Memory leaks can be difficult to  
15 detect, as they slowly degrade machine performance until an out-of-memory error occurs.

Figure 8 is a functional block diagram illustrating one example of a stop code and illustrative self-diagnostic procedure that involves a crash due to allocated pages in  
20 memory not being deallocated. Figure 8 illustrates the above-described out-of-memory error that involves the failure of a driver to properly deallocate its allocated memory when no longer needed, (including when the driver is unloaded). As is known, a process (e.g., of a driver) may request that a block  
25 of virtual address pages be reserved for use by the process (or

even for another process). In that case, the memory manager 201 selects a number of virtual memory addresses to satisfy the request, and reserves those pages to the process. However, if a driver process fails to release virtual addresses that the driver no longer needs, (including when the driver unloads with pages still allocated thereto), then the amount of available virtual memory in the system will slowly decrease until a low-memory situation occurs, leading to a system crash.

To address a system crash of that type, the memory manager 201, upon the next startup, may begin tracking (e.g., via allocation tables 511<sub>1</sub> - 511<sub>n</sub>) each request for and release of virtual memory. The memory manager 201 may track each driver that reserves memory pages, such as by adding an entry in the allocation table each time a device driver 207 issues a request to allocate memory, and by removing the entry if the device driver 207 later deallocates the memory. The entry includes information sufficient to properly identify which driver requested the memory, and may include additional information to aid diagnosis, such as the fields shown in Figure 7, and the like. If the crash reoccurs, the various allocation tables can be examined to identify which driver or drivers (and/or processes thereof) appeared to have requested the allocation of virtual memory without appropriately deallocating those pages. For example, an unloaded driver may still have pages allocated



thereto, or a driver has an inordinate amount of pages allocated thereto.

Another similar type of error, represented by Figure 9, involves a crash due to available Page Table Entries (PTEs) being exhausted. As is known, when a device driver first accesses a virtual address previously allocated thereto, the memory manager 201 creates a page table entry (PTE), such as PTE 927, to map each virtual page (e.g., virtual page 746) to a location in physical memory (e.g., physical page 940).

However, there are a limited number of PTEs that may be created or maintained by the memory manager 201. If a device driver fails to deallocate used memory and thereby cause the release the PTE when no longer needed, the result may be an exhaustion of available PTEs, leading to a system crash.

To address a system crash of that type, the memory manager 201, upon the next startup, may begin tracking (e.g., via a PTE usage table 909) each creation and release of PTEs in a manner similar to that described above for virtual memory pages. The memory manager 201 may track may track the usage of PTEs by each driver, such as by adding an entry in the allocation table each time a device driver 207 causes a PTE to be created, and by removing the entry if the device driver 207 later frees the memory corresponding to the PTE. The entry may include information sufficient to properly identify which driver requested the PTE, and may include additional information to

aid in diagnosis, such as the number of PTEs created, the process of the driver that caused creation of the PTE, the process that called that process, a pointer to the memory descriptor list (MDL), and the like. If the crash reoccurs,  
5 the PTE usage table 909 can be examined to identify which driver or drivers had PTEs in use at the time of the crash. Thus, if a crash is resulting from a lack of available PTEs, the table 909 will likely identify which driver is using an inordinate amount.

10 Figure 10 is a functional block diagram illustrating an example of another stop code and illustrative self-diagnostic procedure that involves a crash due to locked pages in memory not being released. As is known in the art, certain memory pages allocated to a software component may be trimmed to a  
15 pagefile if not actively used. This allows more addressable memory (termed "virtual memory") to be available than the amount of physical memory actually available. For example, as illustrated in Figure 7, virtual memory page 744 and virtual memory page 746 are logically mapped to the same physical  
20 memory location. However, in the illustrated example, virtual memory page 744 was not actively used and was consequently trimmed to the pagefile 705, (as flagged in the PTE therefor), whereby virtual memory page 746 may be actively used.

A driver may request that certain memory pages (e.g.,  
25 virtual memory page 747) be locked in physical memory so that

the information stored in those pages will not be trimmed to the pagefile 705. This is sometimes done by drivers or software components that need to access certain data as quickly as possible, or simply to maintain constant mappings of virtual  
5 addresses to physical addresses. Drivers are supposed to unlock any pages when it is no longer necessary that the pages remain locked. If a driver fails to do so, and/or even worse, continues to lock additional pages as they are used, over time the number of available physical pages of memory decreases  
10 until a low-memory error occurs, leading to a system crash.

In this case, as illustrated in Figure 10, at system start up, the memory manager 201 detects, by the stop code 311, that the type of crash that occurred corresponds to locked pages not being released by a driver. Thus, the memory manager 201 may  
15 initiate a tracking procedure to track pages in memory as they are locked and unlocked. For instance, the memory manager 201 may create and write information to an allocation table 1009 that identifies each request by a driver for locked pages in physical memory. The memory manager 201 may use the allocation  
20 table 709 to track each driver that requests a locked page, such as by adding an entry in the table 1009 for each request, and by removing the entry if the device driver 207 frees the page. The entry includes information sufficient to properly identify which driver locked the page, and may include  
25 additional information to aid diagnosis, such as the same

general information maintained for tracking virtual memory allocations in general, as described above. Alternatively, simple counts of the locked pages of each driver may be maintained, which may help identify an errant driver, although  
5 not necessarily pinpoint the problem within the driver

The memory manager 201 continues to track the locked status of pages in memory until the crash reoccurs (or, perhaps, until turned off through some external mechanism). In this way, if the crash reoccurs, the allocation table 1009  
10 contains an entry for each locked page in memory with sufficient information to identify each driver that has locked the page. In this way, a system administrator or other debugger can analyze the allocation table 1009, for example to identify which driver has an excessive number of locked pages,  
15 thus identifying the offending driver.

#### CORRUPT MEMORY SYSTEM CRASHES

Another category of system errors that lead to crashes is corrupt-memory. One way in which a component can corrupt  
20 memory is related to the way in which pooled system memory is arranged and used. For many reasons, including performance and efficiency, a common pool of system memory is made available to kernel mode components. Pooled memory is allocated by the memory manager 201 as a block, (e.g., in multiples of thirty-  
25 two bytes), with a header (e.g., eight bytes) at the start of

each block. For example, if forty-four bytes of pooled memory are required by a driver, sixty-four are allocated by the memory manager 201, eight for the header, forty-four for the driver, with the remaining twelve unused. Among other  
5 information, the header includes information that tracks the block size. Then, when the memory is deallocated, the memory manager 201 determines whether this block may be coalesced with any adjacent deallocated blocks, to make larger blocks of memory available for future requests. If so, the header  
10 information including the block size is used to coalesce the adjacent blocks.

While this technique is highly efficient in satisfying requests for memory allocations and then recombining deallocated memory, if an errant component, such as driver  
15 1107, writes beyond its allocated memory block, it overwrites the header of the subsequent block. For example, if a driver requests twenty-four bytes, it will receive one thirty-two byte block, eight for the header followed by the requested twenty-four bytes. However, if the driver writes past the twenty-  
20 fourth byte, the driver will corrupt the next header, whereby the memory manager 201 may, for example, later coalesce the next block with an adjacent block even though the next block may be allocated to another component. As can be appreciated, a multitude of errors may result from the corrupted header.

To further complicate matters, the memory manager 201 or the component having the next block allocated to it (or even an entirely different component) will likely appear responsible for the crash. Indeed, this type of error can be very subtle, such as if the header corruption occurs long after the initial deallocation, possibly after many other components have successfully used the same memory location.

Another way in which an errant driver may crash the system is when a driver frees pooled memory allocated thereto, but then later writes to it after the memory has been reallocated to another component, corrupting the other component's information. This may lead to a crash in which the other component appears responsible. Indeed, this post-deallocation writing can be a very subtle error, such as if the erroneous write occurs long after the initial deallocation, possibly after many other components have successfully used the same memory location. Note that such a post-deallocation write may also overwrite a header of another block of pooled memory, e.g., when smaller blocks are later allocated from a deallocated larger block.

To address these type of error, the memory manager 201 may isolate the memory that is provided to a driver in response to a request for pooled memory. In other words, when memory corruption is suspected, the memory manager 201 does not actually give the active drivers (or a subset of the active

drivers) memory from the pool upon request, but instead provides the memory from another special location. In one alternative, a "driver verifier" tool may be used to perform various tests on certain drivers loaded in the system to  
5 identify the offending driver, i.e., the driver verifier component may be turned on to monitor the actions of active drivers that make use of pooled memory. A preferred driver verifier mechanism is described in copending Patent Application Number 09/447,501 entitled "Method And System For Monitoring  
10 And Verifying Software Drivers," filed on November 23, 1999, assigned to the same assignee as the present invention, and the disclosure of which is hereby incorporated by reference.

Figure 11 is a functional block diagram illustrating a special memory structure 1109 that detects memory corruption.  
15 As generally represented in FIG. 11, one of the tests that may be performed is the detection of memory corruption via writing to a section not allocated to the driver, i.e., memory misuse by overruns and underruns. To detect memory corruption, the memory manager 201 allocates driver memory from the special,  
20 non-shared pool 1109, and monitors that pool for incorrect access. More particularly, when a driver-to-be-tested (e.g., 1107) requests pooled memory, the memory allocated to the driver 1107 is placed on a separate page (that is not pageable to disk). To this end, the memory manager 201 returns a  
25 virtual address to the driver 1107 that is the highest possible

address which allows the allocation to fit on the page, in a virtual memory page pool 1109, whereby the allocated memory is aligned with the end of the page. The remainder of the page that precedes the data, if any, is written with random data, which may comprise a recorded pattern or the like to detect underruns.

To detect overruns, the previous page and the next page in the special page pool 1109 are marked inaccessible. Note that this is accomplished via virtual memory management, wherein each virtual address is associated with a page table entry which comprises a physical address to which the virtual address maps, along with bits that control page access. Thus, the surrounding pages are marked "No Access."

Attempts to access memory beyond the allocation buffer (within a page) are immediately detected as an access violation, as such an access is within the subsequent, "No Access" memory page. Note that writing before the beginning of the buffer will (presumably) alter the random data, and when the buffer is freed, this alteration will be detected. In either case, a bug check is issued, whereby the offending driver is identified in keeping with the present invention.

Note that underrun detection may be selected for drivers, such that the allocated memory is instead aligned with the beginning of the page. With this setting, underruns cause an immediate bug check, while overruns (may) cause a bug check



when the memory is freed. In actual implementations, underrun errors tend to occur less often than overrun errors.

Another test of memory misuse that is performed is represented in FIG. 12, whereby when the driver 1107  
5 deallocates memory from the special pool 1109, instead of freeing the virtual address space, the entire section is marked "No access." As a result, any subsequent read or write back to the formerly allocated area is also detected via an access violation. However, because memory space is finite, the system  
10 may need to reuse the special pool 1109 at some time. To this end, the special pool is essentially arranged as a queue with respect to the deallocated pages. More particularly, once deallocated, a page is reused only after cycling through a large number (e.g., 50,000 or more) of allocations to other  
15 areas of the pool 1109. To this end, a pointer or the like to the appropriate virtual address in the pool 1109 may be maintained, e.g., to the address of the virtual next location in which space for allocating and bounding the allocated page is available. Note that a page that is still allocated and its  
20 surrounding "No Access" pages are skipped over such that only deallocated pages are reused.

While the invention is susceptible to various modifications and alternative constructions, certain illustrated embodiments are shown in the drawings and have been  
25 described above in detail. It should be understood, however,

that there is no intention to limit the invention to the  
specific form or forms disclosed, but on the contrary, the  
intention is to cover all modifications, alternative  
constructions, and equivalents falling within the spirit and  
5 scope of the invention.